

# Scientific and Technical Information at INRIA

Anne-Marie Vercoustre

*Direction de l'Information Scientifique et de la  
Communication*

*1<sup>st</sup> September 2005*



INSTITUT NATIONAL  
DE RECHERCHE  
EN INFORMATIQUE  
ET EN AUTOMATIQUE



# Overview

- INRIA missions – some figures
- Scientific Information and Communication
  - challenge and strategy
- New Models for Scientific Communication
- INRIA Open Archive
- Early conclusion

# INRIA

## The French Research Institute for Computer Science and Control

Founded in 1967

### Missions

- Fundamental and applied Research
- Training through research (+ teaching)
- Technology transfer (20 years of enterprise creation)

### Principles

- Research excellence
- Advances in science at the service of economy and society
- Transmitting knowledge and research results to a wide audience

# Some Figures

## Staff

- 3500 persons including 2700 scientists
- 1800 persons on INRIA payroll

## Locations

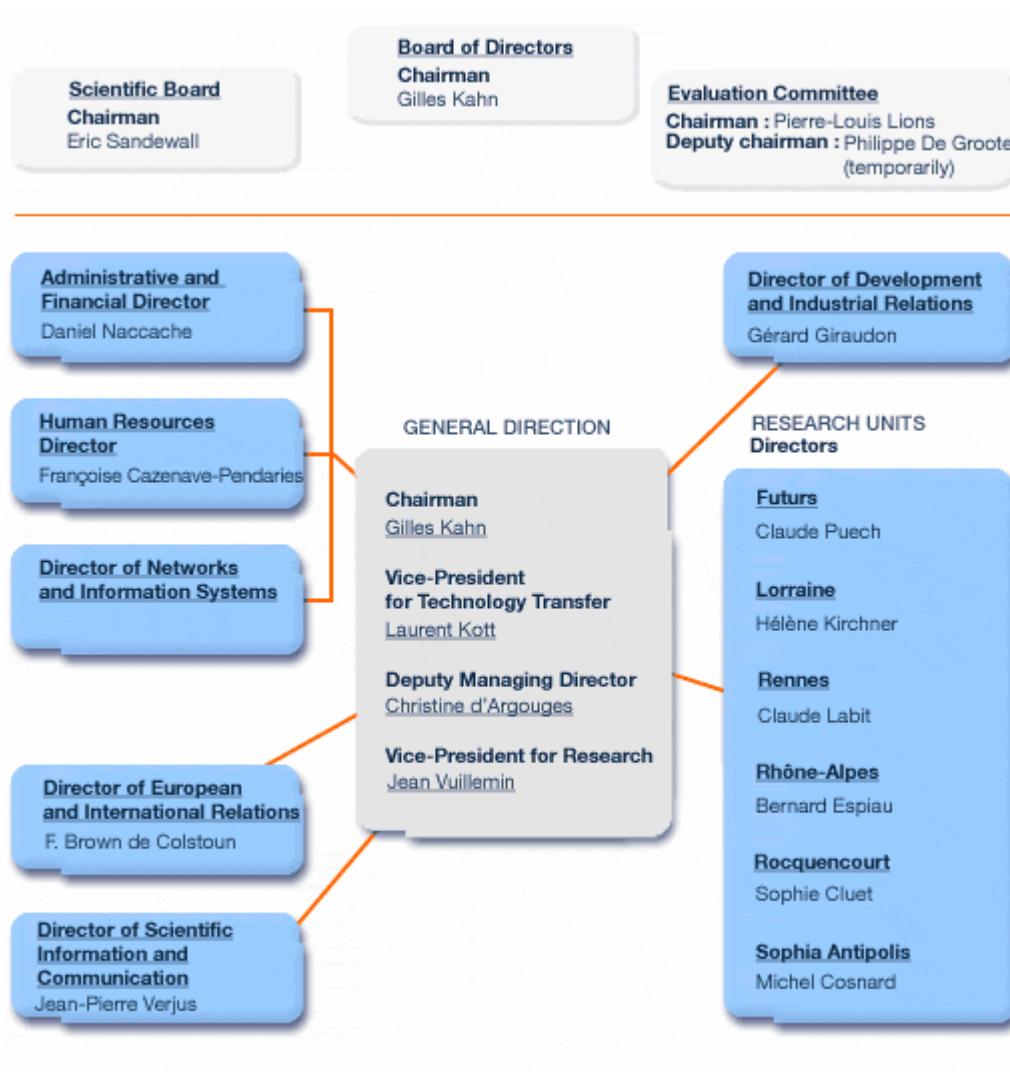
- 6 research units: Rocquencourt, Rennes, Sophia-Antipolis, Lorraine, Grenoble, Futurs (Bordeaux, Lille, Paris)

## Structure

- 146 teams (groups of 10-20 people)
- Five areas of research: communicating systems, cognitive systems, symbolic systems, numerical systems and biological systems

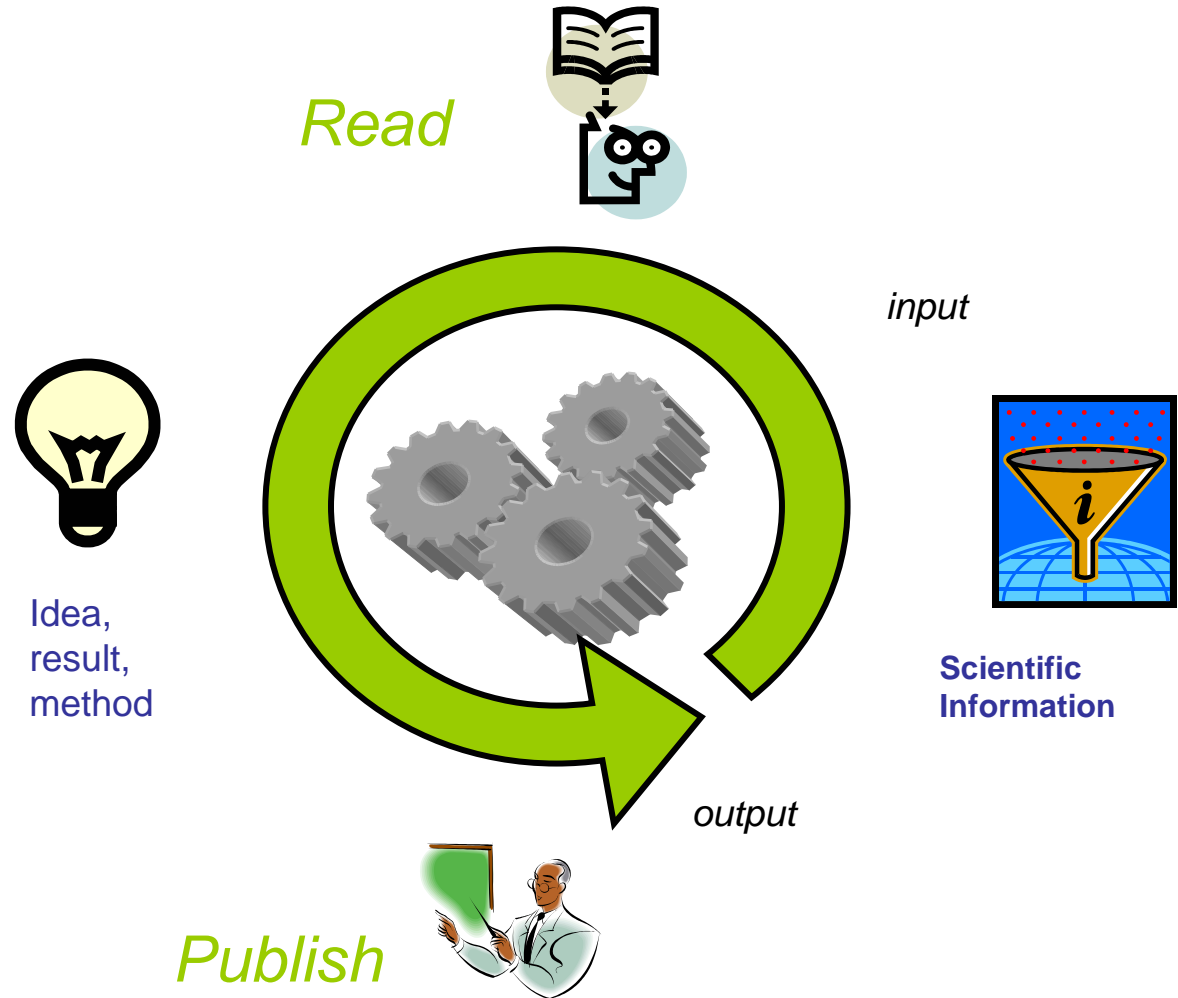
## Budget

- 123 millions Euros, exclusive of tax (2004)
- 22% self-resources



**In Charge of Training through Research**  
Jean-Pierre Verjus

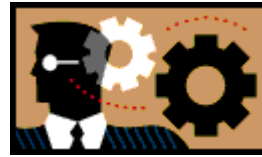
# Scientific Communication



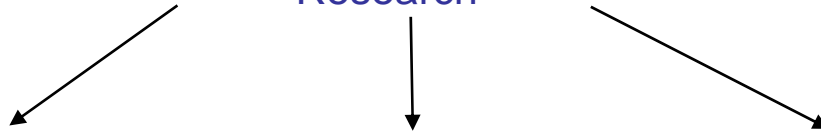
*... a vital process for research*

# Scientific Communication ...

... one form of research valorisation,



Research



Scientific Communication



Technology Transfer



Start-up creation

*Impact can be measured in term of scientific excellence, notoriety, attractiveness.*

# Scientific communication ...



... managed by a specific INRIA service since 2001:

team « IST » (Scientific and Technical Information)

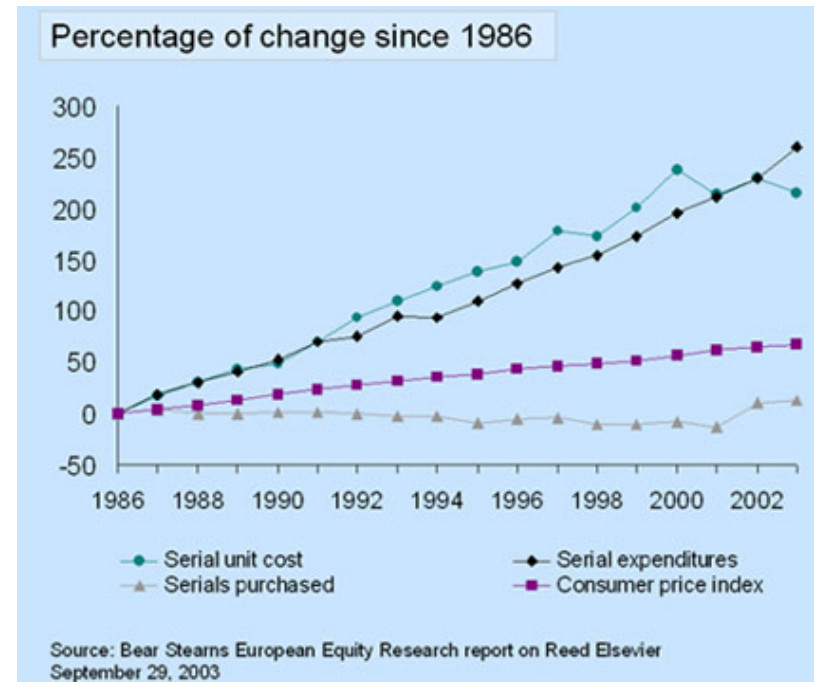
- **Global Budget in 2003** : ~1 M Euros (not including salaries)  
< 1% INRIA budget
- **Staff 2004** : 34 persons (31 in Documentation Centres and Research Units)
  - **A large corpus** : 50000 monographs, 15000 thesis,  
60000 research reports, 800 scientific journals
  - **Internal users are distributed and heterogeneous**: 2700 scientists  
(140 teams) in more than 18 locations, including 450 scientists and  
teachers from partner organisations (universities, CNRS etc.)



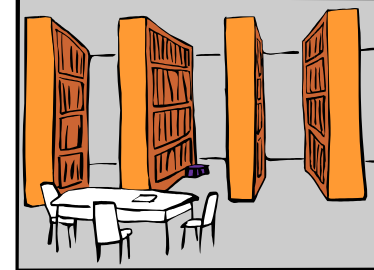
# Access to Scientific literature (1)

## The crisis of publications ...

- For the last 15 years, the cost of publication subscriptions has increased by 215 % and libraries have reduced the number of their subscription by 5%.
- Libraries cannot cope with the need of increasing budgets
- (*Ass. of Research Libraries statistics, 2001*)



# INRIA situation and actions



**Optimising the choice of collections, in priority scientific collections, sharing subscriptions, develop partnerships, negotiating with editors**

**Subscriptions to scientific journals : 80 % of IST budget in 2002**

double, even triple subscriptions...(1100 subscriptions for 660 titles en 2002)

Increasing cost of subscriptions (~+10% per year)

editors' « package », not always appropriate

- A National commission decides subscriptions for the following year
  - 905 subscriptions in 2004 for 600 titles, including 2/3 digital
  - 2003 Budget: 805 kE, e.g. 74 % IST budget, 13% less than the previous year.
- INRIA joined the SUDOC, the network of French University Libraries, to build a common references database.

## Deploy an infrastructure and offer up-to-date services

More and more distributed users

Distributed resources between INRIA libraries

Heterogeneous sites (geographic distances, partners)

heterogeneous access to Information (databases, Editors' portals,

Web portals, search Engine, Open Archives, ...)

- **A common access to serials: Eurydice**

- Collections located in Rocquencourt, Rennes, Sophia, Loria
- Article scanned where it is located
- Article printed where it is needed (displayed in degraded mode)



- **A new Library software (Millenium, Innovative in.): an integrated and efficient management tool for libraries, and an unique INRIA catalogue**

### *Medium Term Objective:*

A common and flexible library system accessed through a dedicated portal



# Publication of Scientific information

**Collecting , adding value, spreading out and archiving INRIA scientific production («grey literature» and publications)**

## Scientists' production

- 400 INRIA research reports in 2003
- 3000 publications (thesis, HDR, communications, articles, monographs, rapports, etc.)
- Activity reports (150 reports a year)
- Free Software (about 50 software on CD-ROM)

## Communication to broader audiences

- **Inedit**, magazine for INRIA scientific partners
- **Bleu INRIA**, a newsletter for elected representatives and public decision makers
- Multimedia scientific presentations
- **Interstice**, a Web site dedicated to « discovery » in Information and Communication Science and technology

# Publication of Scientific information

## New Challenges:

- How to make it accessible to the world?
- How to add value and increase impact?
- How to measure scientific performance?
- Are there new types of data to consider?
- How to archive it in the long term?

# New Models of Scientific Communication (1)

International and strong movement for:

## *Ownership of scientific communication by scientists*

- Ideological claim : free access to knowledge; Open Archive Initiative (OAI)
- Pragmatic concerns : Libraries budgets cannot cope with increasing Publishers' prices

## *New economic models for scientific publishing*

- Separation between different functions in the traditional process :  
reviewing – archiving – diffusion – recording – indexing ...
- Redistribution of roles between different actors (authors, publishers)
- Direct scientific communication

# New Models of Scientific Communication (2)

## What about the Publishers?

- Many are revisiting their publishing policies and copyrights agreements.

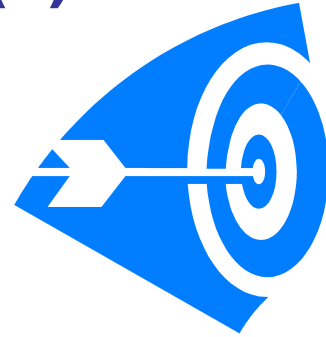
Example : Elsevier

« An author may post his version of the final paper on his personal web site and on his institution's web site (including his institutional repository). Each posting should include the article's citation and link to the journal homepage (or the article DOI). The author does not need our permission to do this, but any other posting (e.g. to a repository elsewhere) would require our permission. By his version we are referring to his Word or TeX file, not a pdf or HTML download from science direct, but the author can update his version to reflect changes made during the refereeing and editing process.... »

*K. Hunter senior vice president strategy Elsevier, may 2004*

- Other possible models (cost paid by the authors)

# New Models of Scientific Communication (3)



Impact of Free access Journals: first study (ISI).

- Out of 8700 Journals ranked by Web of Science about 200 free access Journals have qualified successfully for the ISI
- No significant difference in impact for the two types of publications (fees/free of charge). Given the short history of free journals this is encouraging.
- ISI will follow up on this study

# New Models of Scientific Communication (4)

- CiteSeer is compatible with the OAI protocol and can be searched by OAI search engines (« oaister »)
- Google Scholar (beta test), November 2004
  - **Free Access to scientific articles**
  - **Citations**
- Citations increased by 2.5 to 7 for online publications
  - « *On line or invisible* » ? S. Lawrence, NEC Research Institute, (Nature 2001)
- Who gets acknowledged : measuring scientific contributions through automatic acknowledgement indexing (Proc. of the Nat Academy of Sciences, Giles and Council, Dec 21, 2004)



# L'INRIA engaged in « Open Access »

*Signature of the Berlin Declaration (July 2004)*

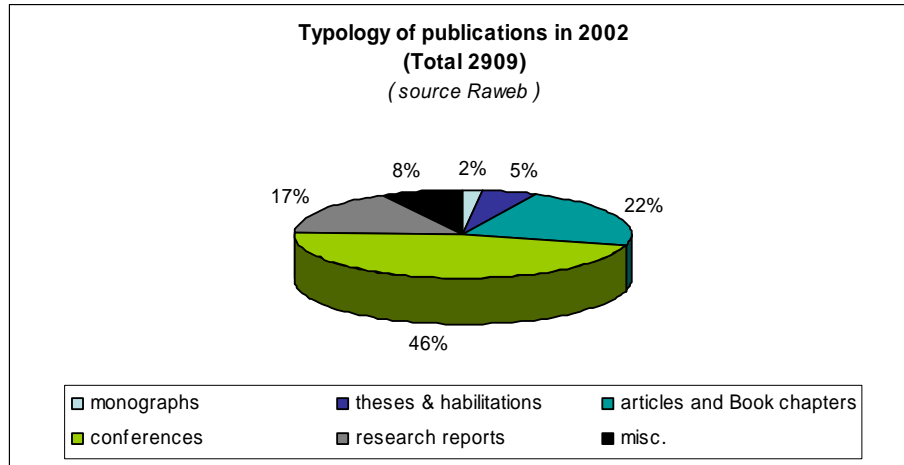
*Formal agreement CNRS-INRIA about direct scientific communication, (July 2004)*

*Collaboration on a software framework for self-archiving:  
HAL (Hypertext Archive on Line)*

*hosted by CCSD (Lyon)*

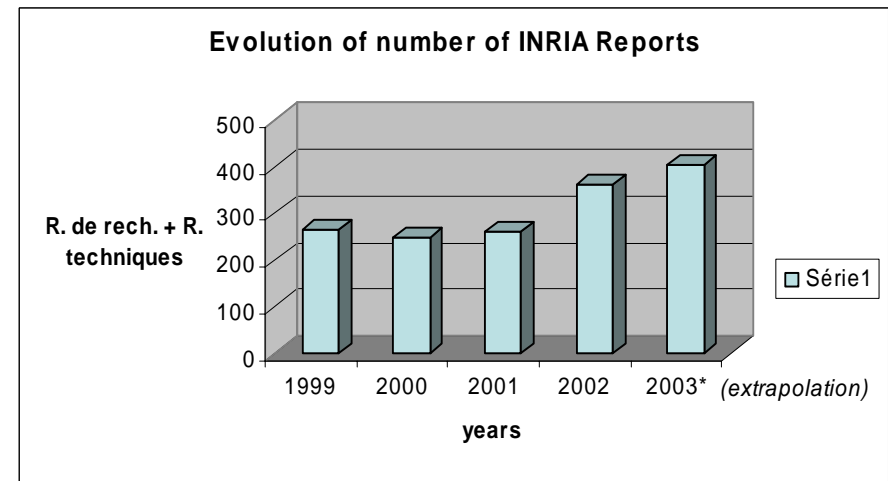


# INRIA Publications



Importance of  
Conference publications

Research Reports: an INRIA  
publication



# Combining Open Archive and Institutional Archive (1)

- Self-archiving by INRIA scientists, not mandatory; no scientific moderation
- Librarians' assistance for checking/completing metadata; technical moderation
- Visible/invisible in accordance to copyrights
- Optional cross-posting to ArXiv
- Educating scientists on juridical aspects of traditional publishing; promoting self-archiving (preprint, postprint)
- Archiving full Conferences proceedings

# Combining Open Archive and Institutional Archive (2)

- Specific Interface for INRIA staff (non exclusive)
- Recording authors' laboratory to keep track of INRIA/team/other labs/ publications
- Automatic cross-posting of INRIA Reports (retroactively)
- Annual Activity Reports maintain an exhaustive list of annual INRIA publications (mostly without the full text articles)
- Improve the production and exploitation process of the Activity Report to make it a portal to INRIA scientific activities (indicators, references and content)

# INRIA Activity Reports

- A mandatory annual report by every team (for the last 20 years)
- In English since 2003
- Predefined but flexible format : *members, objectives, foundation, new results, software, industrial contracts, International and national cooperation, scientific community, publications*
- Authored in Latex or XML;
- Automatic generation into HTML (Web and CD-ROM) and XML (Xyleme database)

## Project Ralyx

Taking advantage of XML for semi-automatic exploitation  
(Views and indicators)

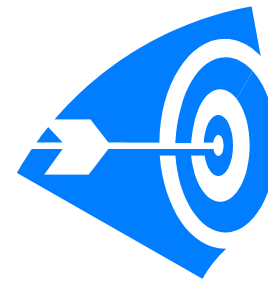
# HAL INRIA opened in April 2005

- Well received by Scientists (with a lot of questions)
- Slow taking off (133 articles in 4 months)
- Several demands for conference proceedings (INRIA scientist editor)

## Open issues:

- Better understanding of the new publication models and process  
(preprint; preprint ->postprint; versions; cross-posting)
- Retroactive posting of INRIA reports
- Thesis, thematic archives
- Software (beyond description and patent)
- Data (e.g. used in experiments)

# Conclusions



Scientific communication is a strategic activity for a Research Organisation

A fast evolving context requires a prospective vision and strategy for INRIA. We embrace the Open Archive vision and hope that CNRS and INRIA together can demonstrate its benefits to the French community.

Activities and pilot projects are local to INRIA Units; actors being the Documentation Centres, and the Scientists who are also the users.

National strategy and coordination is structuring the overall effort and actions.

Alliance with partners is paramount (CNRS, other EPST, Europe,...)